

Felles infrastruktur og tjenester for FAIR forskningsdata

Rapport fra spørreundersøkelsen til forskerne sommeren 2021

Publisert 15.03.2022

Utarbeidet av: Sekretariatet for utredningen

Kontakt: katrine.weisteen.bjerde@hkdir.no



Dette verk er lisensiert under en Creative Commons Navngivelse 4.0 Internasjonal lisens.

Innholdsfortegnelse

1. Bakgrunn.....	3
1.1. Gjennomføring	3
1.2. Vurdering	3
2. Resultater	4
2.1. Bakgrunnsinformasjon	4
2.1.1. Institusjon og fagområde.....	4
2.1.2. Stilling, arbeidsform og ansvar	5
2.1.3. Datatype, format og volum	7
2.2. Planlegging av datahåndtering.....	9
2.2.1. Retningslinjer og krav	9
2.2.2. Tjenester for planlegging av datahåndtering	10
2.2.3. Øvrige veiledningstjenester brukt i planleggingsfasen.....	11
2.2.4. Behov i planleggingsfasen som ikke dekkes av tjenester per i dag	12
2.3. Håndtering av data i pågående prosjekter	13
2.3.1. Lagringsmedium	14
2.3.2. Lagringstjenester	14
2.3.3. Tjenester for analyse og bearbeiding	15
2.3.4. Behov i den aktive fasen som ikke dekkes av tjenestene per i dag	16
2.4. Datahåndtering ved avslutning av prosjekter.....	17
2.4.1. Beslutninger om bevaring av data.....	17
2.4.2. Tjenester for arkivering av data.....	17
2.4.3. Behov i avslutningsfasen som ikke dekkes av tjenestene per i dag.....	19
Vedlegg 1: Oversikt over tjenester	21
Vedlegg 2: Oversikt over spørsmål.....	24

1. BAKGRUNN

Utredningen «Felles infrastruktur og tjenester for FAIR forskningsdata» (videre kalt FAIR-utredningen) gjennomførte i 2021/2022 to kartlegginger. En spørreundersøkelse sendt til forskere og en spørreundersøkelse sendt til institusjonene. I dette notatet oppsummeres funnene fra spørreundersøkelsen til forskerne.

1.1. GJENNOMFØRING

Undersøkelsen ble sendt ut på epost til postmottak ved institusjoner i UH-sektoren, institutter som er medlemmer av FFA¹ og helseforetakene, adressert til forskningsledelsen. Institusjonene ble bedt om å sende ut undersøkelsen til sine forskere, men vi har ikke oversikt over hvem som gjorde det og hvor mange de eventuelt sendte til. Vi kjenner derfor ikke til svarprosenten på undersøkelsen. Spørreundersøkelsen lå også åpent tilgjengelig på openscience.no. Det var to versjoner av undersøkelsen, en norsk og en engelsk. Se oversikt over spørsmålene i vedlegg 2.

1.2. VURDERING

Det var totalt 531 forskere som besvarte spørreundersøkelsen. Det er en liten andel av forskerne i Norge. Vi opplever likevel relativt god spredning i deltakere med tanke på erfaring med datadeling og type data de håndterer. Alle fagområder er representert, men det er noe overvekt på medisin og helsefag.

Undersøkelsen har derfor ikke blitt brukt til å trekke konklusjoner om større grupper forskere, men har vært et nyttig utgangspunkt for å se på overordnede trekk og finne temaer til videre arbeid i utredningen.

¹ <https://www.abelia.no/bransjeforeninger/ffa-forskningsinstituttene-fellesarena/medlemmer/>

2. RESULTATER

I dette kapitlet presenteres resultatene fra spørreundersøkelsen. Undersøkelsen var delt i fire hoveddeler:

- Bakgrunnsinformasjon om deltakerne
- Planlegging av datahåndtering
- Håndtering av data i pågående prosjekter
- Datahåndtering ved avslutning av prosjekter

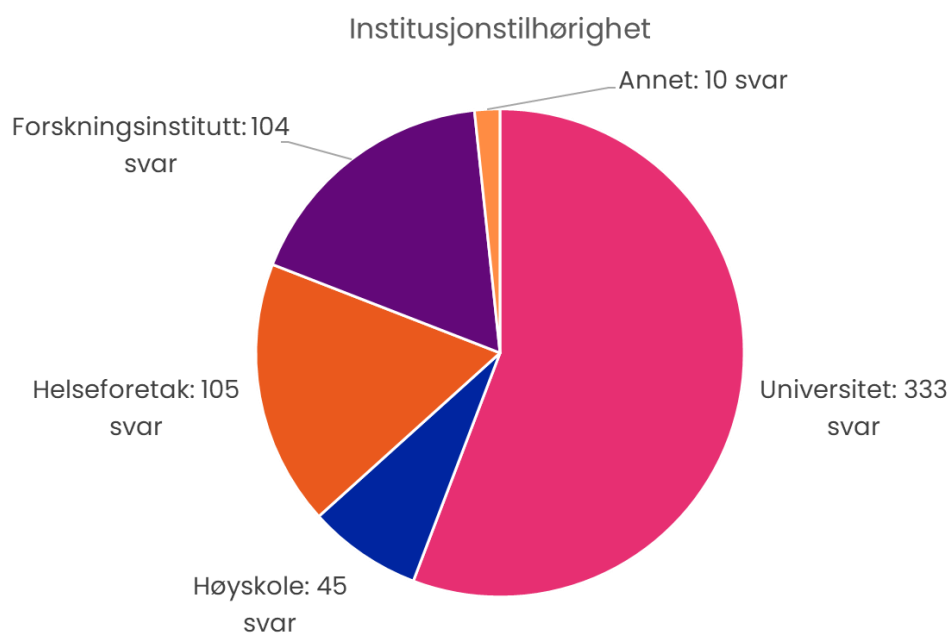
2.1. BAKGRUNNSINFORMASJON

I undersøkelsens første del stilte vi spørsmål om deltakernes institusjonstilhørighet, fagbakgrunn og stilling, samt spørsmål om arbeidsform og dataene de behandlet. Hensikten med spørsmålene var å bruke informasjonen som bakgrunnsvariabler for analyse av de andre spørsmålene i undersøkelsen.

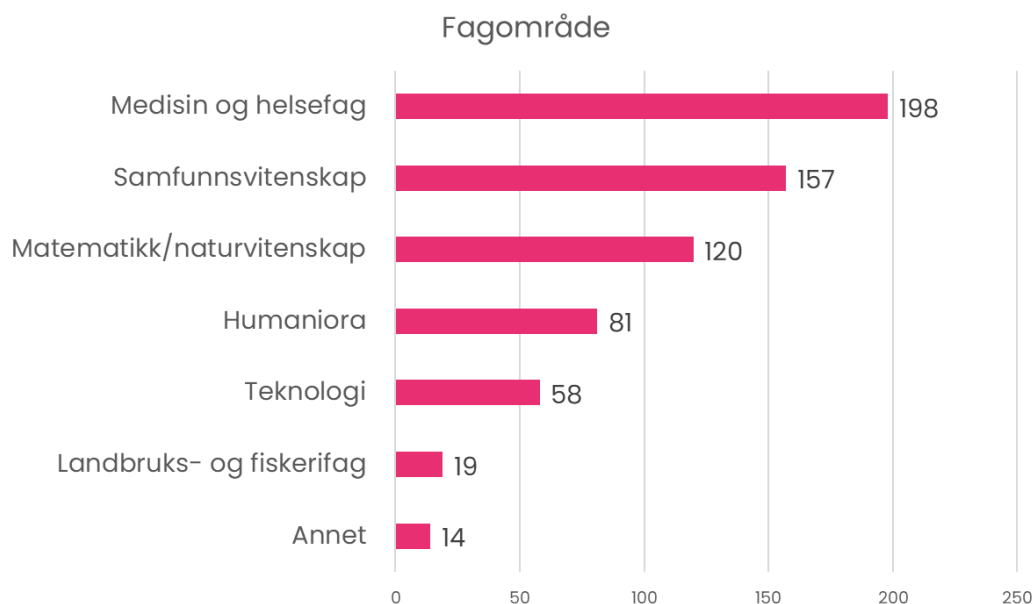
2.1.1. Institusjon og fagområde

For å kartlegge deltakernes bakgrunn stilte vi spørsmålene «Hvilken type institusjon er du tilknyttet?» og «Hvilke(t) fagområde jobber du innen?». Det var mulig å krysse av flere alternativer på begge spørsmål.

De fleste deltakerne kom fra universitetene, etterfulgt av helseforetakene og forskningsinstituttene. Det største fagområdet var medisin og helsefag, etterfulgt av samfunnsvitenskap.



Figur 1: Oversikt over deltakernes institusjonstilhørighet (antall svar = 597)

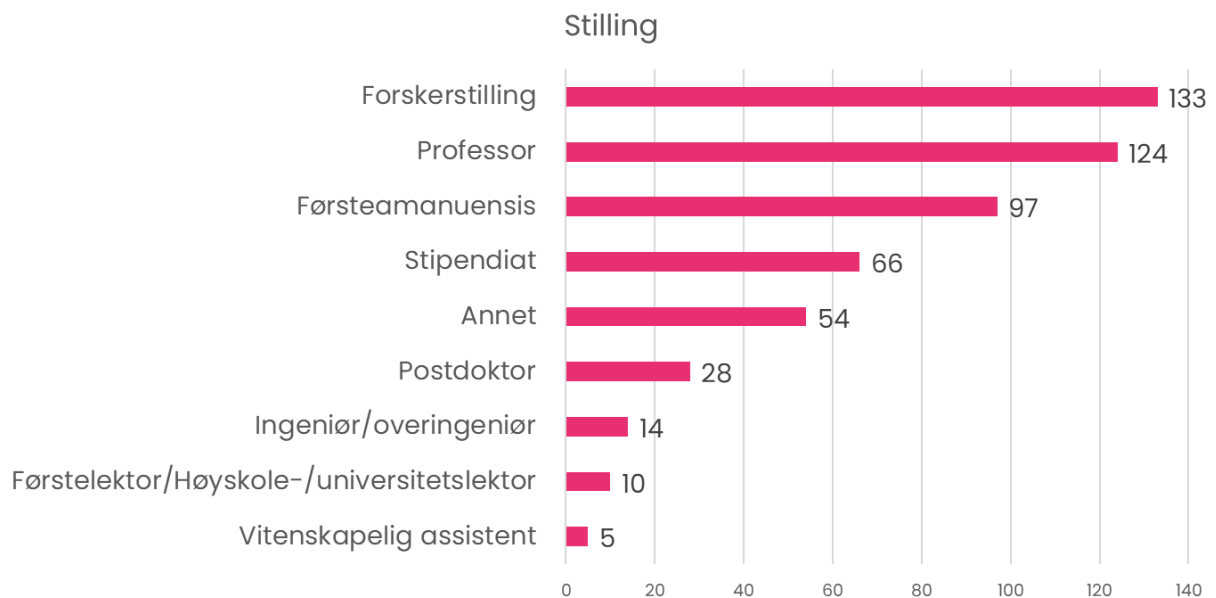


Figur 2: Oversikt over deltakernes fagområde (antall svar = 647)

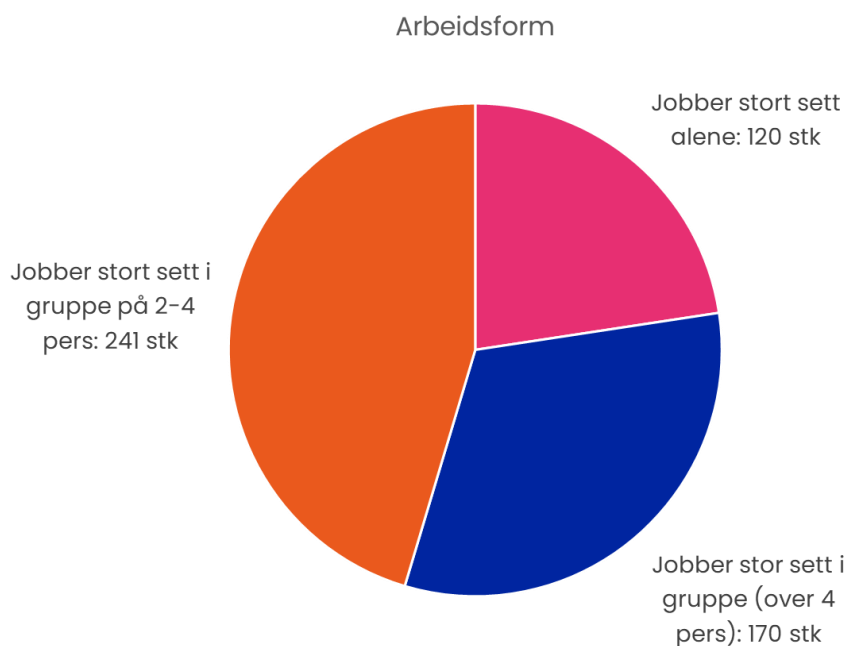
2.1.2. Stilling, arbeidsform og ansvar

For å undersøke deltakernes stilling, arbeidsform og ansvar stilte vi spørsmålene: «Hvilken stilling er du ansatt i?», «Jobber du i forskningsgrupper eller jobber du stort sett alene?» og «Er du vanligvis prosjektleder/PI eller deltaker i prosjektene du deltar i?». Det var kun mulig å velge ett alternativ på disse spørsmålene.

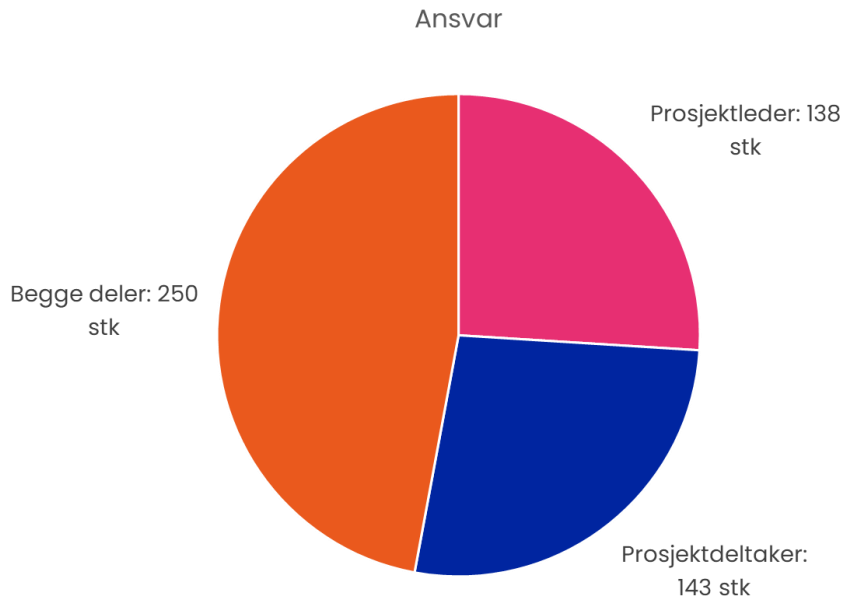
De største stillingskategoriene var forskerstilling, professor og førsteamanuensis. 77 % arbeider stort sett i grupper på 2-4 stk. eller flere. 73 % er prosjektledere enten vanligvis eller av og til. Det ser dermed ut til at undersøkelsen har truffet forskere med forskererfaring, behov for å samarbeide underveis i prosjekter og som har erfaring med oppgaver knyttet til prosjektledelse.



Figur 3: Oversikt over stillingene deltakerne er ansatt i (antall svar = 531)



Figur 4: Oversikt over deltakernes arbeidsform og samarbeid (antall svar = 531)

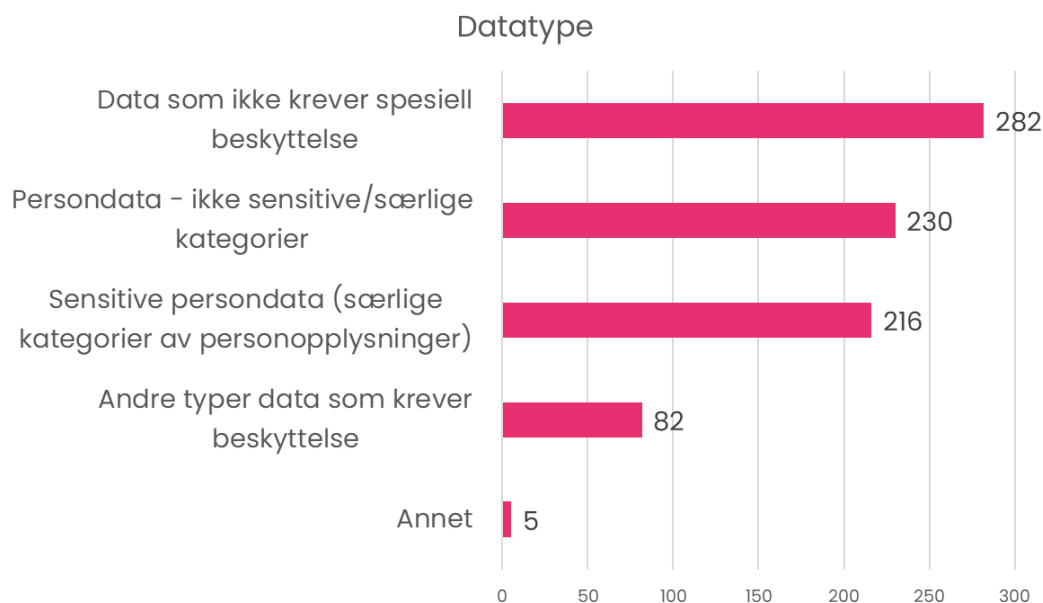


Figur 5: Oversikt over deltakernes prosjektansvar (antall svar = 531)

2.1.3. Datatype, format og volum

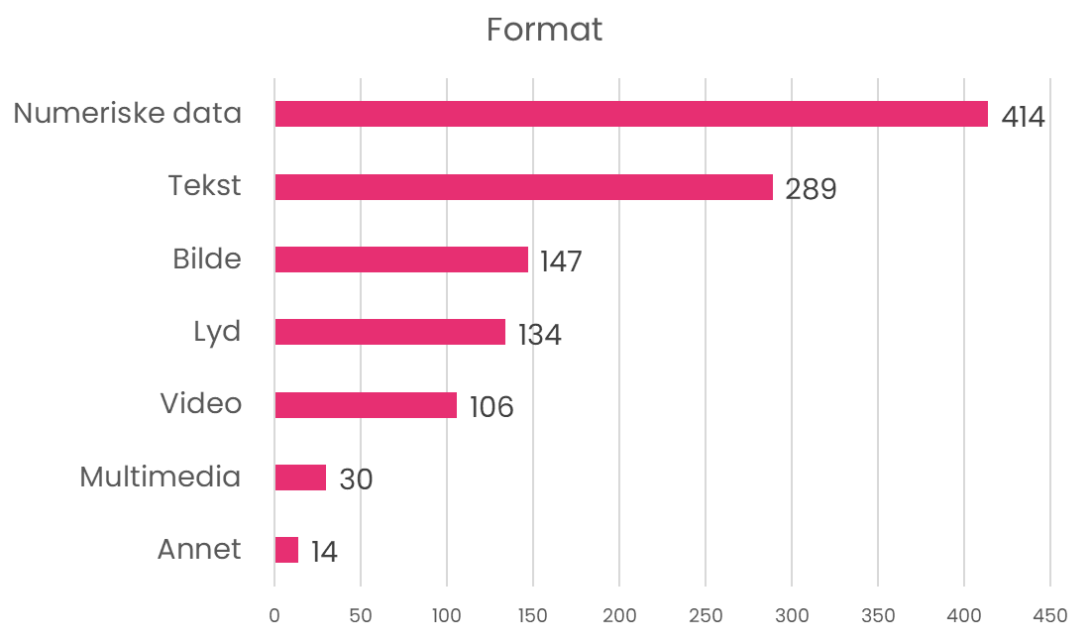
For å få mer innsikt i dataene deltakerne håndterte stilte vi spørsmålene: «Hvilke(n) datatype(r) håndterer du primært i dine prosjekter?», «Hvordan foreligger dataene du håndterer i dine prosjekter?» og «Hvor store datavolum håndterer du normalt i dine prosjekter (per prosjekt)?». Det var mulig å krysse av flere alternativer på alle spørsmålene.

Det var god spredning i svarene på spørsmålene om type, format og volum. Det høye antallet svar viser også at mange håndterer flere datatyper og formater i sine prosjekter.



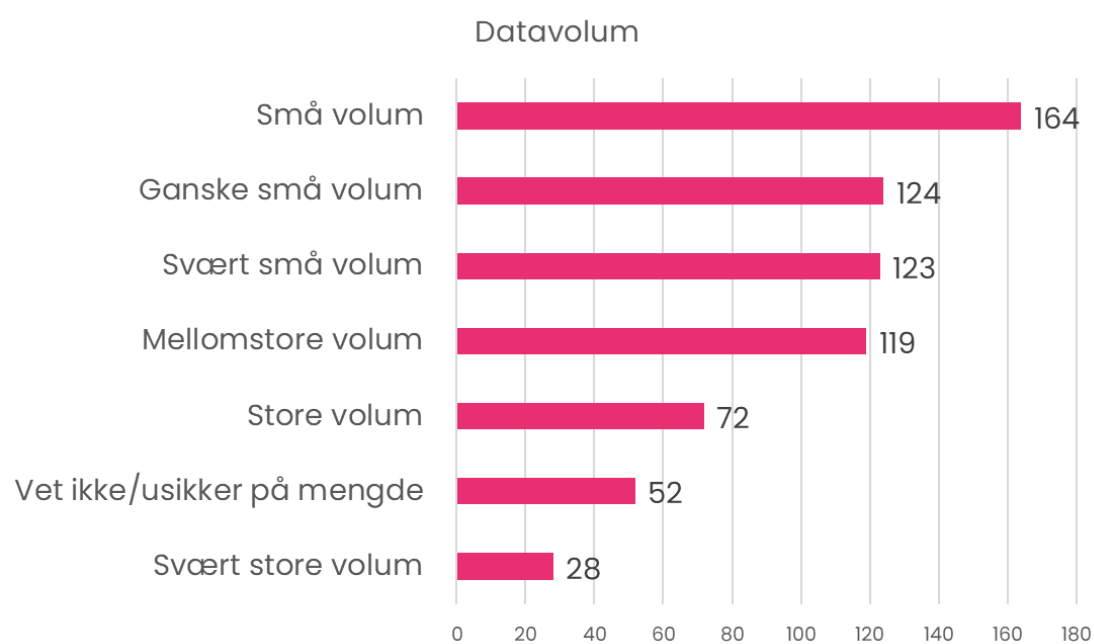
Figur 6: Oversikt over datatypene deltakerne håndterer (antall svar = 815)

Eksempler på datatyper som ble nevnt eller spesifisert i fritekstfeltet var data med opphavsrettsbeskyttelse.



Figur 7: Oversikt over formatet på dataene deltakerne håndterer (antall svar = 1134). Forklaringer som er forkortet: Numeriske data (databaser, regneark, målinger, laboratoriedata, simuleringer etc.) og Tekst (dokumenter)

Eksempler på dataformater som ble nevnt eller spesifisert i fritekstfeltet var DNA-sekvenseringsdata og geografiske data.



Figur 8: Oversikt over datavolumet deltakerne vanligvis håndterer i sine prosjekter (antall svar = 682). Se vedlegg med detaljert oversikt over spørsmål for definisjoner av de ulike kategoriene.

2.2. PLANLEGGING AV DATAHÅNDBTERING

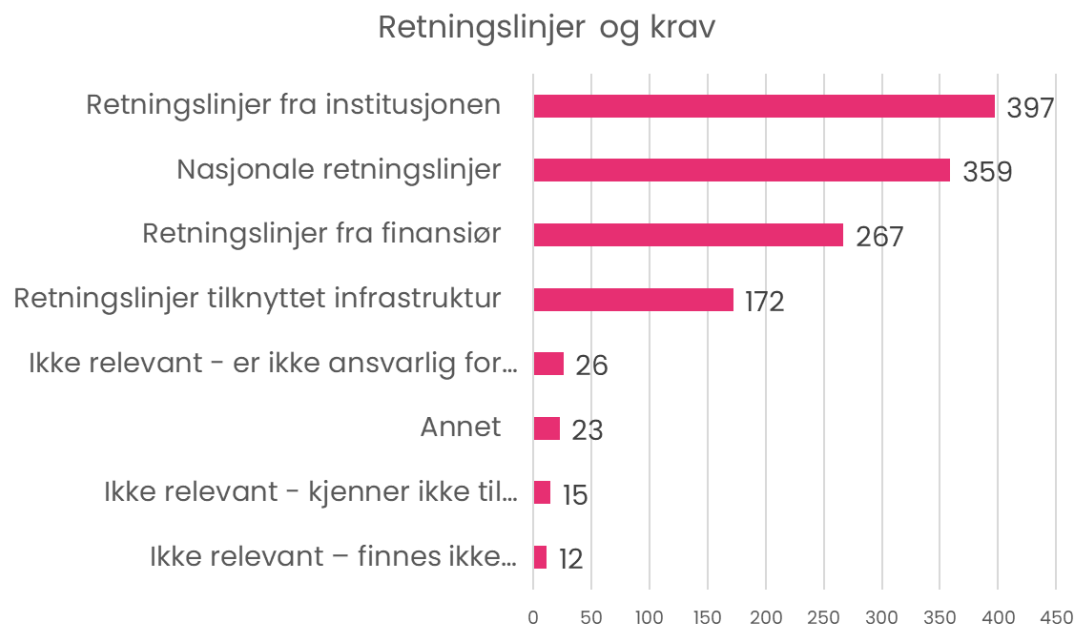
I undersøkelsens andre del stilte vi spørsmål knyttet til prosjektplanleggingsfasen. Spørsmålene handlet om hvilke retningslinjer forskerne forholdt seg til, hvilke tjenester de brukte for å planlegge datahåndtering, hvilke øvrige veiledningstjenester de hadde benyttet og om de hadde andre behov som ikke ble dekket av tjenestene i dag.

2.2.1. Retningslinjer og krav

Bolken innledet med spørsmålet: «Hvilke retningslinjer og krav til datahåndtering forholder du deg til når du planlegger dine forskningsprosjekter?»

Det store antallet svar på spørsmålet viser at forskerne forholder seg til flere ulike retningslinjer og krav i planleggingen av sine prosjekter. Tilnærmet 75 % av deltakerne forholder seg til retningslinjer fra egen institusjon, som viser betydningen av institusjonenes retningslinjer for deltakerne i vår undersøkelse.

Av de 26 som svarer at retningslinjene ikke er relevant fordi de ikke er ansvarlige for planlegging av datahåndtering, er det kun to stk. som også har krysset av at det ikke er relevant fordi de ikke kjenner til retningslinjer eller at det ikke finnes retningslinjer. Det ser dermed ikke ut til å være overlapp her.



Figur 9: Oversikt over hvilke retningslinjer og krav deltakerne forholder seg til i planleggingen av sine prosjekter (antall svar = 1271). Forkortelser i oversikten: Ikke relevant – finnes ikke egne retningslinjer på mitt område, Ikke relevant - kjenner ikke til retningslinjer på området, Ikke relevant - er ikke ansvarlig for planlegging av datahåndtering. Se også oversikt over spørsmål for detaljerte svaralternativer.

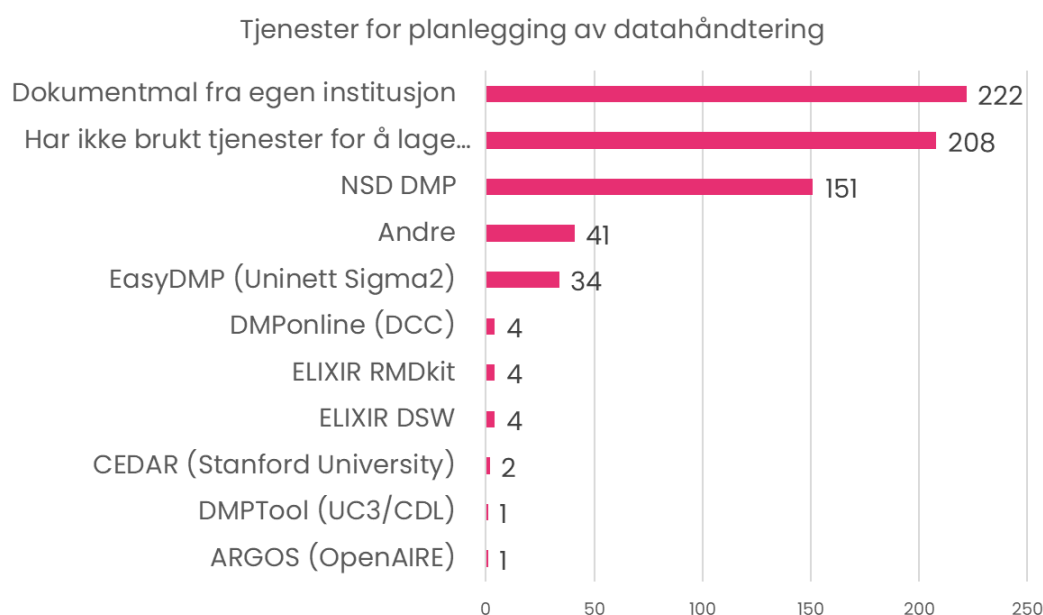
Eksempler på retningslinjer som ble nevnt eller spesifisert i fritekstfeltet var retningslinjer fra de forskningsetiske komitéer eller NSD.

2.2.2. Tjenester for planlegging av datahåndtering

Videre stilte vi spørsmål om hvilke tjenester forskerne har brukt/bruker under planleggingen av datahåndtering i sine forskningsprosjekter. Det var mulig å krysse av flere alternativer.

Det var flest som hadde erfaring med dokumentmal fra egen institusjon og NSDs DMP. Det var også en relativt stor gruppe som ikke hadde erfaring med tjenestene.

19 av de som ikke har ansvar for planlegging har ikke brukt tjenester.



Figur 10: Oversikt over bruk av tjenester for planlegging av datahåndtering (antall svar = 672)

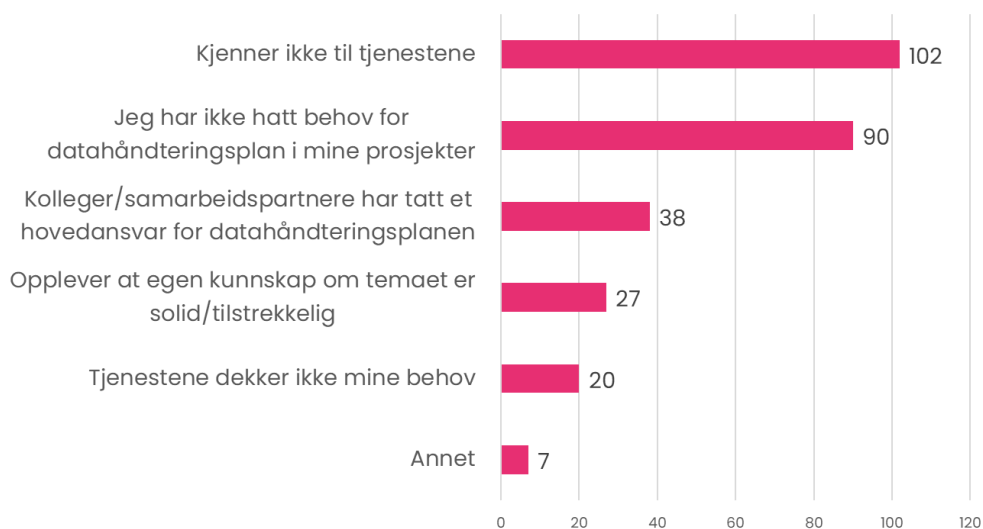
For oversikt over andre tjenester for planlegging av datahåndtering nevnt av deltakerne, se vedlegg 1.

Årsaker til manglende bruk

Deltakerne som krysset av at de ikke hadde benyttet tjenester fikk oppfølgingsspørsmål om hva som var årsaken til at de ikke har benyttet tjenester for å lage datahåndteringsplan. Det var mulig å krysse av flere alternativer.

Hovedårsaken oppgitt av deltakerne var manglende kjennskap til tjenestene. Det var en noe større andel av deltakerne fra universitetene og helseforetakene som oppga at de ikke hadde kjennskap til tjenestene. Det var også en større andel av de som ikke hadde prosjektledererfaring som ikke kjente til tjenestene. Men det var ingen signifikante forskjeller. En del oppgir også at de ikke har hatt behov for datahåndteringsplan eller at noen andre har tatt ansvar for planen.

Årsaker til manglende bruk av tjenester for planlegging

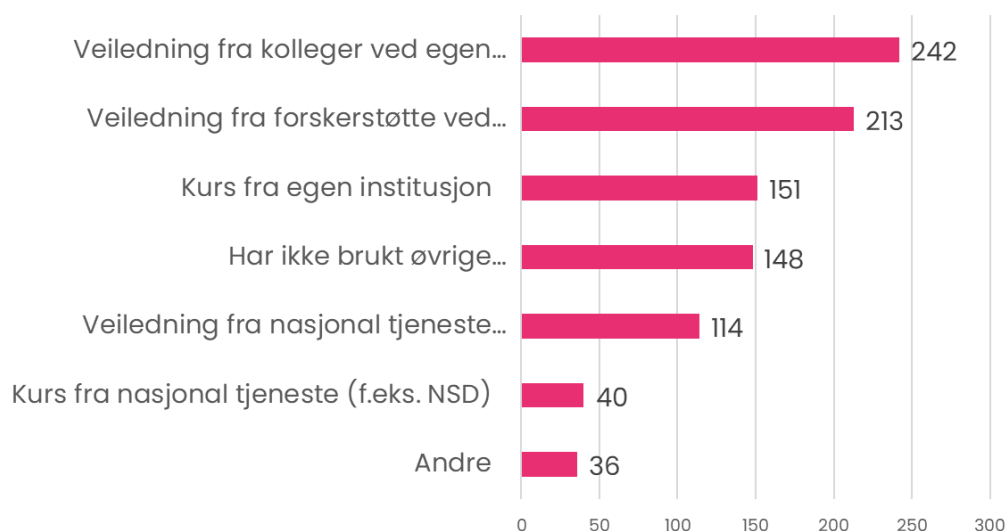


Figur 11: Årsaker til manglende bruk av tjenester for planlegging (antall svar = 284)

2.2.3. Øvrige veiledningstjenester brukt i planleggingsfasen

I tillegg til konkrete tjenester for planlegging undersøkte vi hvilke øvrige veiledningstjenester forskerne har brukt/bruker under planleggingen av datahåndtering i sine forskningsprosjekter. Det var mulig å krysse av flere alternativer.

Øvrige veiledningstjenester



Figur 12: Forskernes erfaring med øvrige veiledningstjenester i planleggingsfasen (antall svar = 944). Forkortelser: ved egen ... = ved egen institusjon. Se detaljert oversikt over spørsmål for informasjon om svaralternativene.

Eksempler på øvrige veiledningstjenester fra fritekstsvarene:

- Mal fra annen institusjon
- Veiledning fra kolleger ved annen institusjon

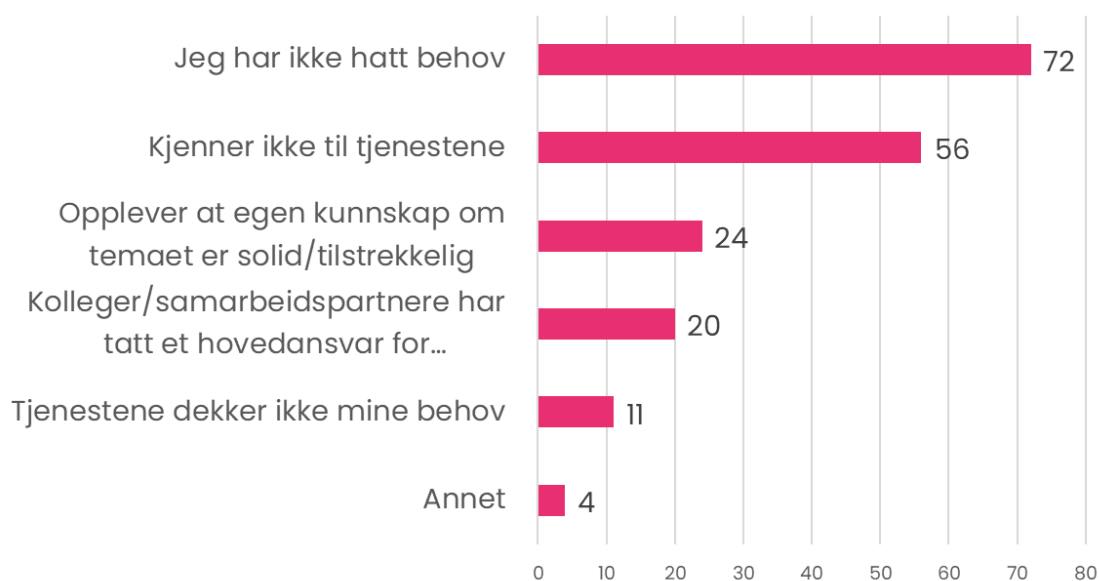
- Veiledning fra finansjør
- Veiledning fra regional forskningsstøtte

Årsaker til manglende bruk

Deltakerne som krysset av at de ikke hadde benyttet tjenester fikk oppfølgingsspørsmål om hva som var årsaken til at de ikke hadde benyttet øvrige veiledningstjenester. Det var mulig å krysse av flere alternativer.

Hovedårsaken til manglende bruk var at forskerne ikke hadde hatt behov for veiledning, etterfulgt av manglende kjennskap til tjenestene. Det var en noe større andel av deltakerne fra universitetene og helseforetakene som oppga at de ikke hadde kjennskap til tjenestene. Det var også en større andel av de som ikke hadde prosjektledererfaring som ikke kjente til tjenestene. Men det var ingen signifikante forskjeller.

Årsaker til manglende bruk av veiledningstjenester

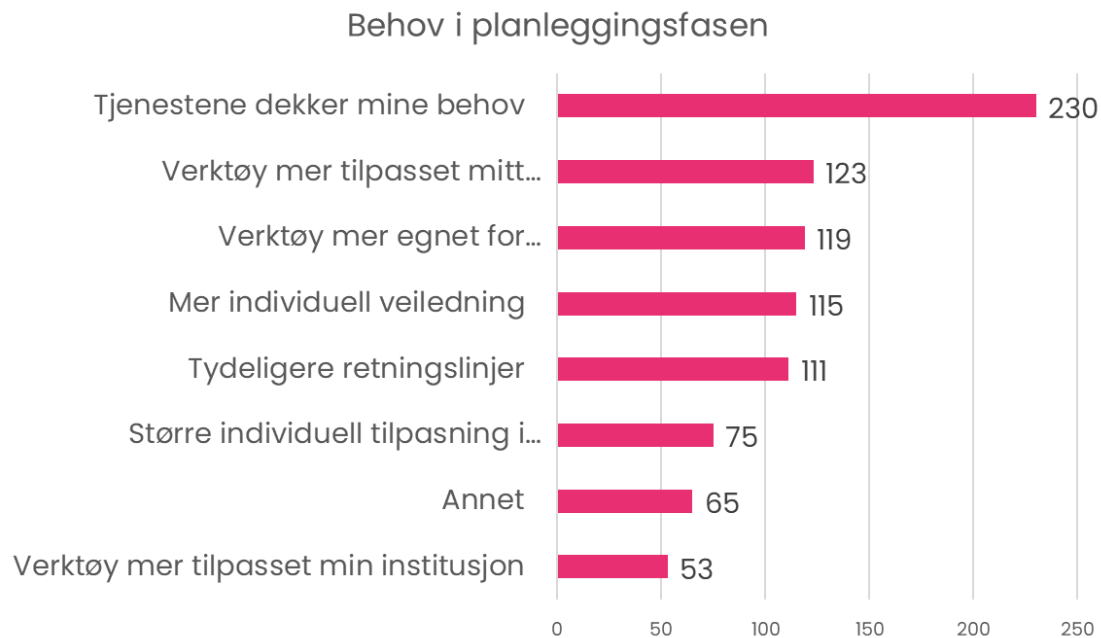


Figur 13: Årsaker til manglende bruk av øvrige veiledningstjenester (antall svar = 187)

2.2.4. Behov i planleggingsfasen som ikke dekkes av tjenester per i dag

Til sist stilte vi spørsmål om forskerne hadde behov knyttet til planlegging av datahåndtering som ikke dekkes av tjenestene per i dag. Det var mulig å krysse av flere alternativer.

Flest svarte at tjenestene dekker deres behov. Av behov som ikke dekkes per i dag var det flere som ønsket seg verktøy mer tilpasset deres fagområde og de forskningsmetodene de brukte. Det var også et ønske om mer individuell veiledning og tydeligere retningslinjer.



Figur 14: Behov som ikke dekkes av tjenestene per i dag (antall svar = 891). Forkortelser: Behov for større individuell tilpasning i verktøyene, Behov for verktøy mer egnet for forskningsmetodene jeg bruker, Behov for verktøy mer tilpasset mitt fagområde, Behov for større individuell tilpasning i verktøyene

Eksempler på fritekstsvar:

- Institusjonen overlater i dag dette i for grad til forsker selv å håndtere - savner en (sic.) institusjonell fokus på datahåndtering
- Ja! Jeg jobber som kvalitativ forsker. Et stort behov jeg har er det som kalles "active citation" internasjonalt, der man dokumenterer utvalgte deler av det kvalitative kildematerialet. En slik tjeneste forutsetter at man kan vise fram de kildene som ikke er sensitive, for eksempel en skriftlig kilde som er delvis anonymisert, eller deler av et intervju i sin hele lengde. Det ideelle er hvis man samtidig kan lagre sensitive data i samme tjeneste, dvs det som ikke kan deles offentlig. Noe slikt finnes ikke i Norge i dag.
- Jeg har behov for å forstå hvordan jeg kan tilpasse prosjektet mitt til regelverket og praktiske løsninger. Min opplevelse er at f.eks. (institusjon) sine "guidelines" er vanskelige å forstå, språk er unødvendig vanskelig og jeg trenger hjelp til å "oversette" dette inn i hvordan jeg faktisk skal løse dette i praksis. Her har NSD vært utrolig behjelpelig, men de har f.eks. begrenset informasjon om ulike fillagringsmuligheter ved (institusjon).

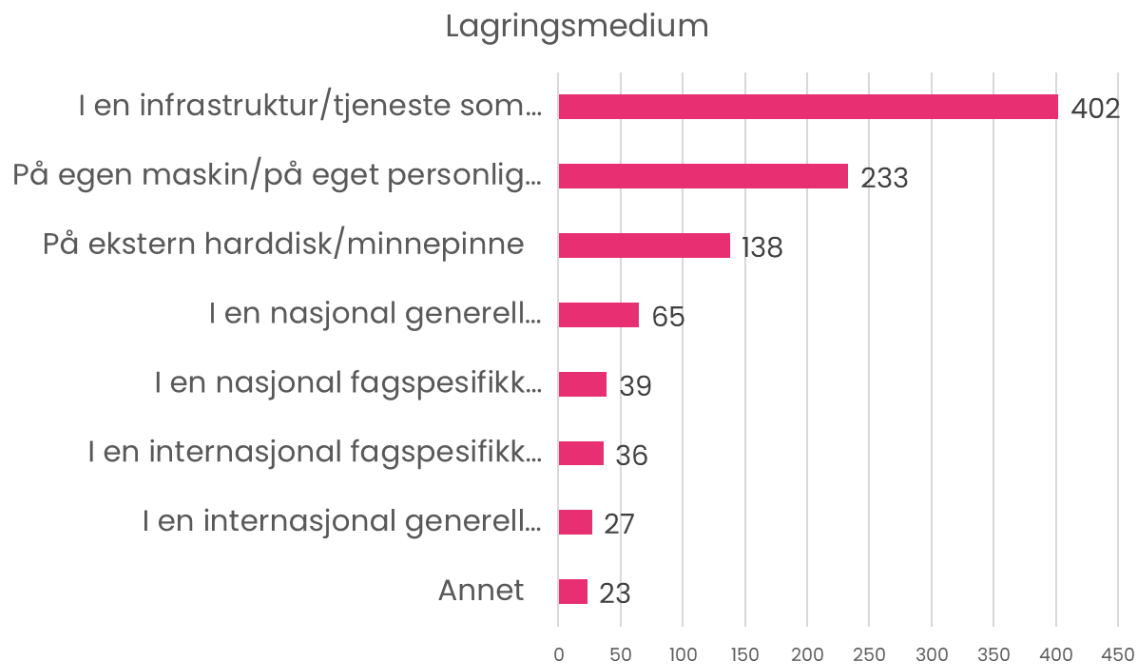
2.3. HÅNTERING AV DATA I PÅGÅENDE PROSJEKTER

I undersøkelsens tredje del stilte vi spørsmål om den aktive fasen av forskningsprosjektene. Vi undersøkte hvor forskerne primært lagret data, hvilke tjenester de brukte for lagring og analyse av aktive data og om de hadde noen behov i den aktive fasen som ikke ble dekket av tjenestene i dag.

2.3.1. Lagringsmedium

Innledningsvis stilte vi spørsmål om hvor forskerne primært lagrer data i sine pågående forskningsprosjekter. Det var mulig å velge flere alternativer.

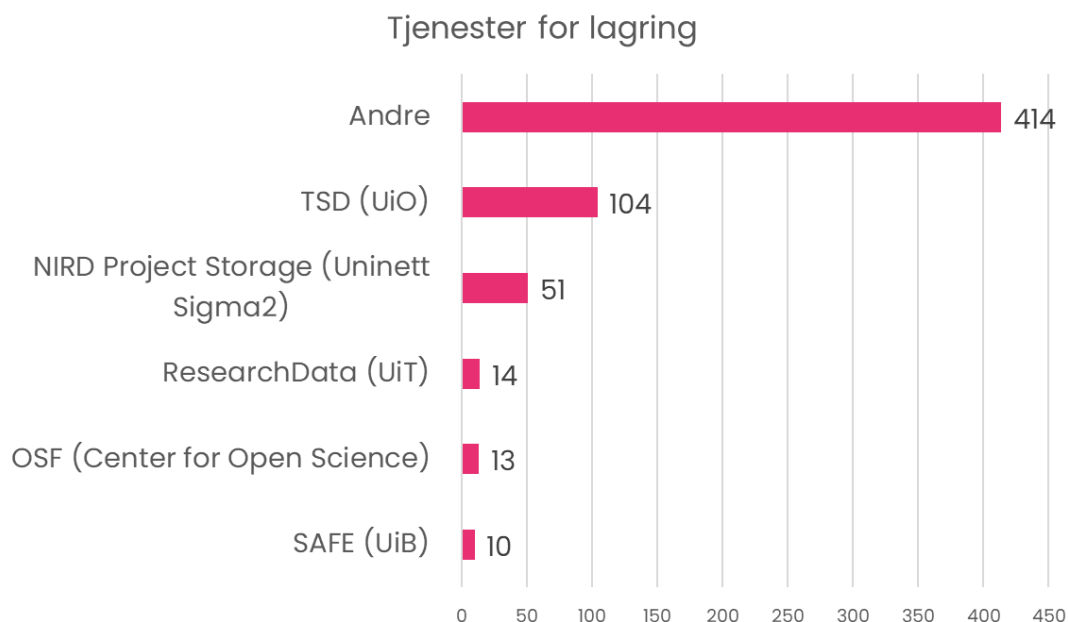
De fleste lagrer data på en tjeneste/infrastruktur som tilhører institusjonen. En del lagrer også på egen maskin og ekstern harddisk/minnepinne.



Figur 15: Oversikt over hvor forskerne primært lagrer data i pågående prosjekter (antall svar = 936). Forkortelse i oversikten: I en infrastruktur/tjeneste som tilhører institusjonen

2.3.2. Lagringstjenester

Videre stilte vi spørsmål om hvilke(n) tjeneste(r) forskerne har brukt/bruker for å lagre data i den aktive fasen av forskningsprosjekter. Det var mulig å krysse av flere alternativer.

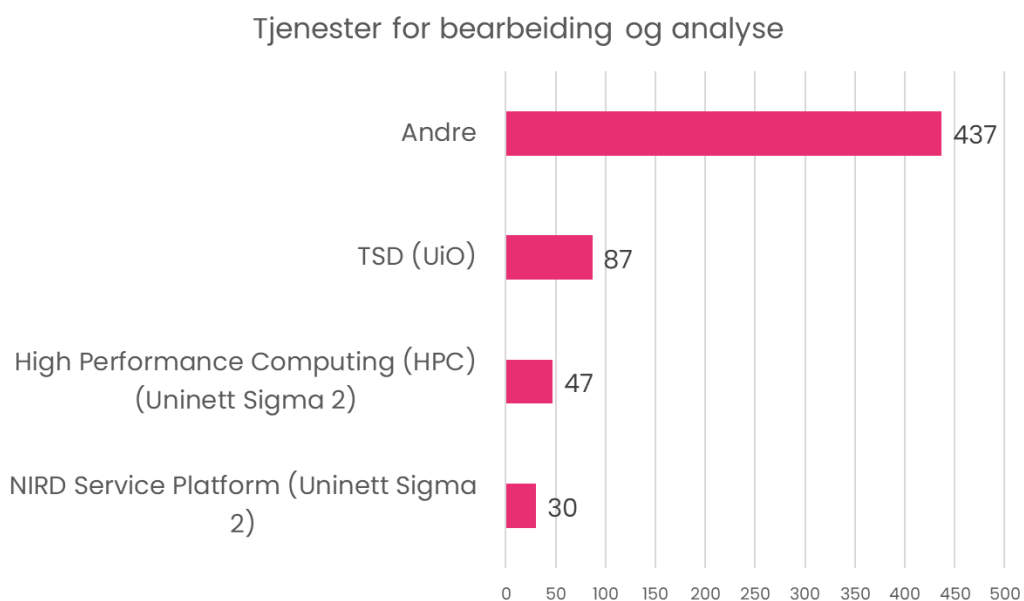


Figur 16: Oversikt over bruk av tjenester for lagring av aktive data (antall svar = 606)

Av andre tjenester var det 67 stk. som nevnte lokal server/tjeneste fra egen institusjon. Se vedlegg 1 for oversikt over andre tjenester nevnt i fritekstsvarene.

2.3.3. Tjenester for analyse og bearbeiding

Vi stilte også spørsmål om hvilke(n) tjeneste(r) forskerne har brukt/bruker for å analysere og bearbeide data i den aktive fasen av forskningsprosjekter. Det var mulig å krysse av flere alternativer.



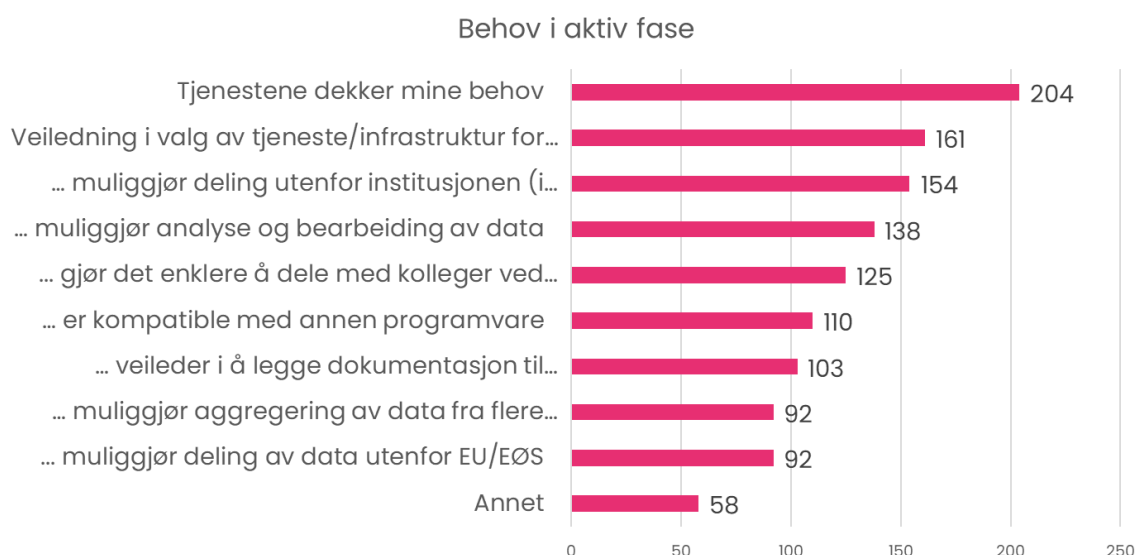
Figur 17: Oversikt over bruk av tjenester for bearbeiding og analyse av data (antall svar = 601)

Eksempler på tjenester nevnt i fritekstsvarene var Nvivo, HUNT cloud, Excel og SPSS. For oversikt over andre tjenester, se vedlegg 1.

2.3.4. Behov i den aktive fasen som ikke dekkes av tjenestene per i dag

Til slutt stilte vi spørsmål om forskerne har behov knyttet til datahåndtering i den aktive fasen som ikke dekkes av tjenestene per i dag. Det var mulig å krysse av flere alternativer.

Flest svarer at tjenestene dekker deres behov. Mange ønsker seg veiledning i valg av tjeneste/infrastruktur – hva som er egnet når. Slår vi sammen noen av alternativene ser vi at tjenester som muliggjør deling (både internt, innen Norge/EØS og utenfor EØS) er en stor gruppe.



Figur 18: Oversikt over behov i aktiv fase (antall svar = 1237)

Forkortelser i figuren:

- Tjenester som muliggjør deling av data i den aktive fasen med samarbeidspartnere utenfor EU/EØS (definert som «tredjeland» i GDPR)
- Tjenester som muliggjør aggregering av data fra flere kilder
- Tjenester som veileder i å legge dokumentasjon til dataene (f.eks. legge på gode metadata)
- Tjenester som er kompatible med annen programvare jeg bruker
- Tjenester som gjør det enklere å dele data i den aktive fasen med kolleger ved institusjonen
- Tjenester som muliggjør analyse og bearbeiding av data
- Tjenester som muliggjør deling av data i den aktive fasen med samarbeidspartnere utenfor institusjonen (i Norge og innen EU/EØS)
- Veiledning i valg av tjeneste/infrastruktur for lagring av data - hva kan brukes når?

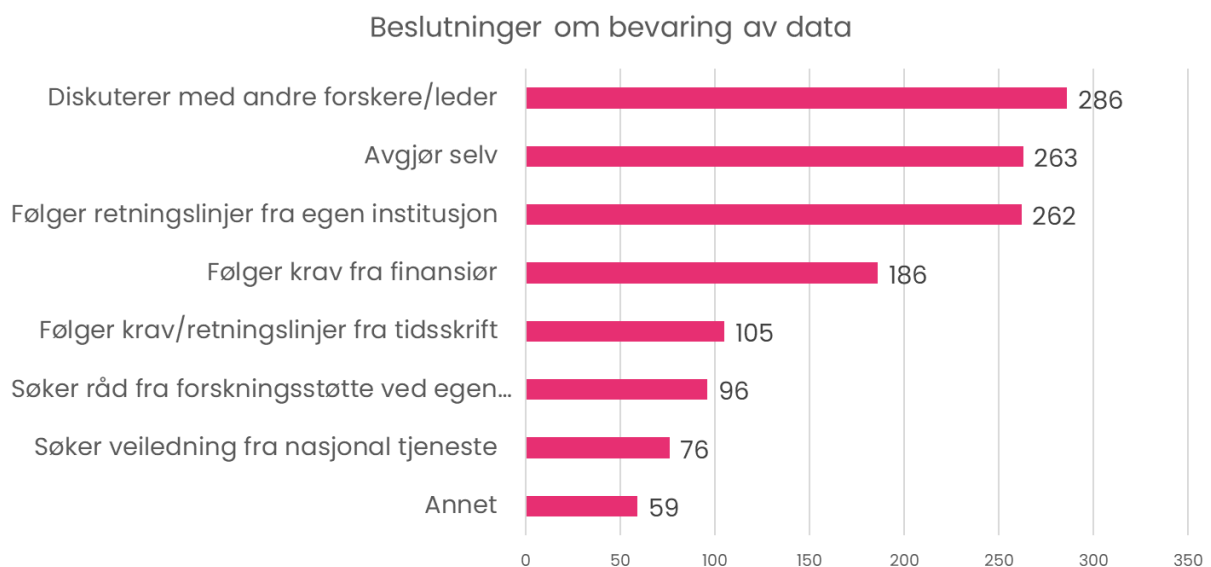
2.4. DATAHÅNTERING VED AVSLUTNING AV PROSJEKTER

I undersøkelsens siste del stilte vi spørsmål om prosjektavslutningsfasen. Vi undersøkte hvordan forskerne avgjør hvilke data som skal spares på etter prosjektets slutt, hvilke tjenester for arkivering de bruker og om de hadde behov i avslutningsfasen som ikke dekkes av tjenesten per i dag.

2.4.1. Beslutninger om bevaring av data

Innledningsvis stilte vi spørsmål om hvordan forskerne går frem for å avgjøre hvilke data de skal ta vare på etter prosjektets slutt. Det var mulig å velge flere alternativer.

Her var det mange som krysset av flere alternativer og det ser ut til at mange henter råd fra ulike kilder før de avgjør.



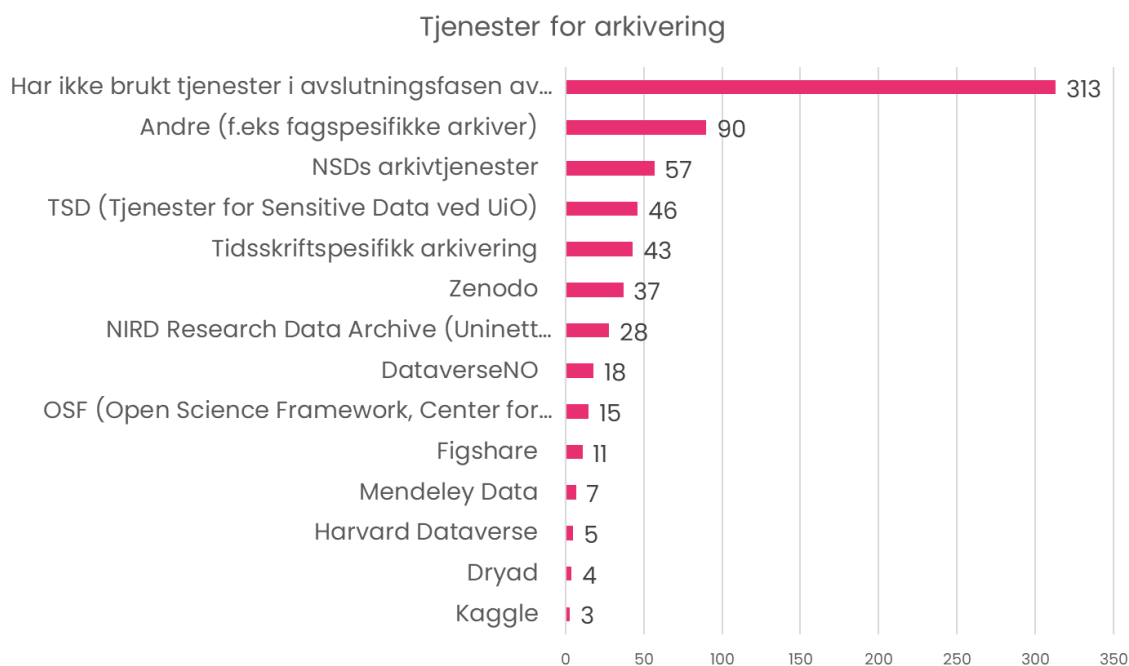
Figur 19: Oversikt over hvordan forskerne går frem for å beslutte hvilke data som skal bevares etter prosjektslutt (antall svar = 1333)

2.4.2. Tjenester for arkivering av data

Videre stilte vi spørsmål om hvilke av de følgende tjenestene forskerne har brukt/bruker for å arkivere eller dele data i avslutningsfasen/etter avsluttet forskningsprosjekt. Det var mulig å velge flere alternativer.

Her er det mange som ikke har brukt tjenester for arkivering. Ser vi på de detaljerte svarene så er det få som har krysset av både for at de ikke har brukt og for konkrete tjenester. Det ser dermed ikke ut til at mange har valgt alternativet fordi de i tidligere prosjekter ikke har arkivert. Av de som har erfaring med konkrete tjenester er den største gruppen andre, f.eks. fagspesifikke arkiver, etterfulgt av NSDs arkivtjenester.

Ser vi videre på hvem som har svart at de ikke har erfaring med arkivering er det en større andel fra høyskolene og helseforetakene som ikke har brukt tjenester for arkivering. Denne forskjellen var signifikant.



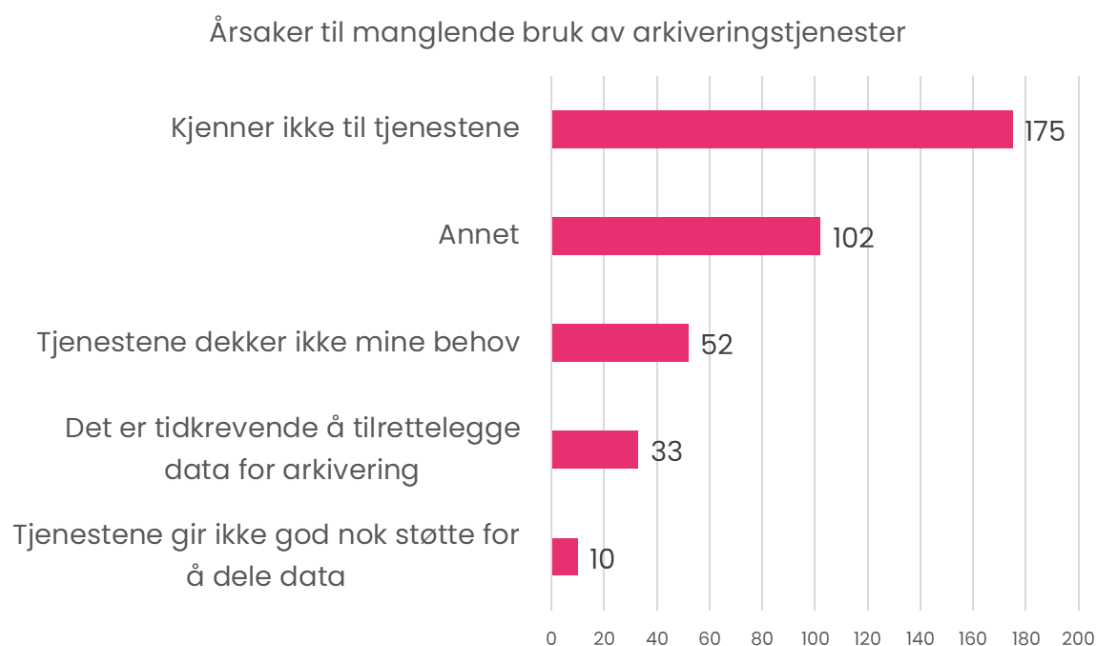
Figur 20: Oversikt over bruk av tjenester for arkivering eller deling av data (antall svar = 677)

Eksempler på tjenester fra fritekstsvarene er EBRAINS, Pangea og PhysioNet. Se vedlegg 1 for flere tjenester.

Årsaker til manglende bruk

Deltakerne som krysset av at de ikke hadde benyttet tjenester fikk oppfølgingsspørsmål om hva som var årsaken til at de ikke hadde benyttet tjenester for arkivering. Det var mulig å krysse av flere alternativer.

Hovedårsaken til manglende bruk oppgis å være manglende kjennskap til tjenestene. Det var en større andel fra høyskolene og helseforetakene som oppga at de ikke hadde kjennskap til tjenestene. Det var også en større andel av de som ikke hadde prosjektledererfaring som ikke kjente til tjenestene. Men det var ingen signifikante forskjeller.

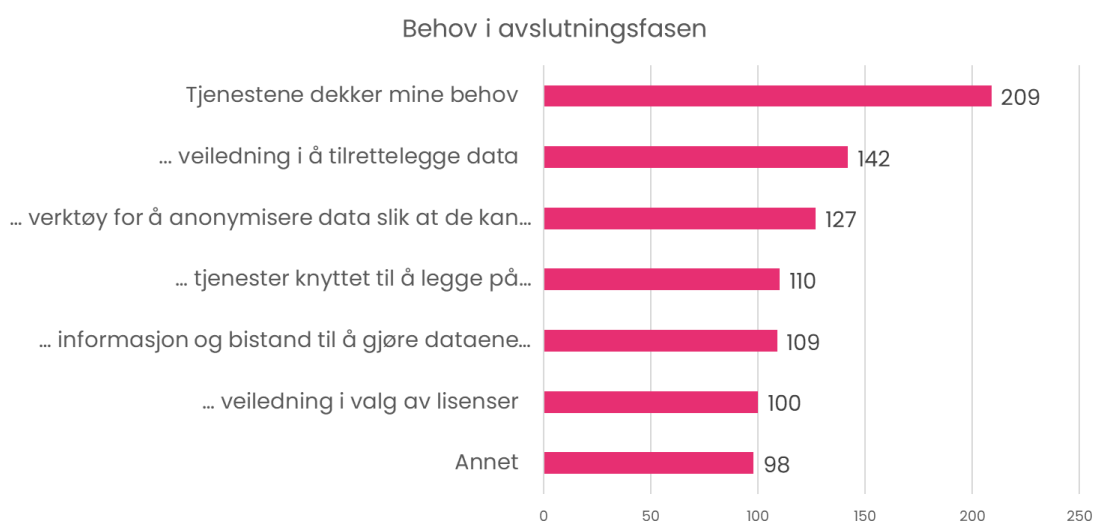


Figur 21: Oversikt over årsaker til manglende bruk av arkiveringstjenester (antall svar = 372)

2.4.3. Behov i avslutningsfasen som ikke dekkes av tjenestene per i dag

Avslutningsvis stilte vi spørsmål om forskerne har behov knyttet til datahåndtering i avslutningsfasen som ikke dekkes av tjenestene per i dag. Det var mulig å velge flere alternativer.

Det er flest som svarer at tjenestene dekker deres behov. Ellers er det svært jevnt mellom de ulike alternativene.



Figur 22: Oversikt over behov i avslutningsfasen (antall svar = 895)

Forkortelser i tabellen:

- Behov for informasjon og bistand til å gjøre dataene siterbare
- Behov for tjenester knyttet til å legge på dokumentasjon/metadata
- Behov for verktøy for å anonymisere data slik at de kan deles
- Behov for veiledning i å tilrettelegge data (valg av filnavn, formater, anonymisering etc.)

Eksempler på fritekstsvar:

- Kjøpte data fra institusjoner (MorBarn undersøkelsen, HUNT, etc) kan ofte ikke deles etter at prosjektet er ferdig.
- Need for spaces to host very large data in any format (e.g. Zenodo has a 50Gb limit).
- Jeg har enda ikke hatt behov for det, men synes det mangler tydelige nasjonale retningslinjer på hva som skal til for at et datasett med personopplysninger blir anonymt nok til at det kan lagres og/eller deles med andre uten at det medfører brudd på taushetsplikten (selv om direkte identifiserbare opplysninger slettes, er det vanskelig å vite i hvilken grad et datasett er indirekte identifiserbart).
- Jeg opplever en nokså merkelig diskrepans mellom det offentliges "push" for åpne, delbare forskningsdatasett og det samtidige enorme fokuset sykehusene har på å etterfølge GDPR-lovgivningen. Der anses ethvert helsedatapunkt som like sensitivt, uansett hvor lite det "sier" om personen (f.eks. min helt normale konsentrasjon av natrium-ioner i blodet en gitt dag, eller at pulsen min er 85 per minutt).

Jeg har hittil ikke sett denne diskrepansen diskutert hverken fra nasjonalt eller internasjonalt hold. Faktisk anonymisering av helsedata (etter personvernombudenes standard) vil oftest ikke være mulig. Videre har jeg aldri sett forslag til tekstutforming av informasjons-/ samtykkeskriv (f.eks fra REK) som ville være dekkende for at en pasient skulle tillate full og ubegrenset deling av deres forskningsdata i all fremtid. Datadeling er jo noe pasientene ville måtte si ja til ved inklusjon i studien, siden de da spesifikt må samtykke til hvordan deres data skal behandles og av hvem

VEDLEGG 1: OVERSIKT OVER TJENESTER

Denne oversikten inneholder tjenester for ulike deler av forskningsprosessen nevnt av deltakerne i spørreundersøkelsen.

Andre tjenester for datahåndteringsplanlegging
Datajoint
DMP fra Oslo universitetssykehus
DMP plan fra INCF norsk node / EBRAINS
DPIA utformet i samarbeid med NSD
Easy Trial
EU Horizon 2020-mal
Mal for datahåndteringsplan fra NORCrins websider for legemiddelstudier (kompatible med EUDRACT retningslinjer)
OpenNeuro
OSF
www.codecheck.org.uk
BioMedData project from ELIXIR
Dokumentmal fra PLOS.
Qualtrics
RCN template.
Science Europe
We have created our own data management standards "OpenMinds" in Ebrains.eu

Andre tjenester for lagring av aktive data hentet fra fritekstsvarene:

BBMRI - BBMRI-ERIC a European research infrastructure for biobanking
Box
Dropbox
Google Drive
GitHub/GitLab
HUNT Cloud
NeLS - Norwegian eInfrastructure for Life Science
NICE1 ved NTNU

Zenodo
SWITCH (Swiss educational consortium)
OneDrive
Qualtrics data base
Wold LHC Computing Grid (WLCG)
AURORA NTNU
RITMO data servers
Infrastrukturen CLARINO og CLARIN ERIC

Tjenester for analyse av data

Nvivo
Excel
AntConc
CLARINO og CLARIN ERIC
R
HUNT Cloud
NICE, NTNU
SPSS
Stata
Medinsight

Tjenester for arkivering

Tjeneste	Beskrivelse	Link
CLARINO	Common Language Resources and Technology Infrastructure Norway	https://clarin.w.uib.no/
CLARIN	Common Language Resources and Technology Infrastructure	https://www.clarin.eu/
EBRAINS		https://ebrains.eu/
PANGAEA	Data Publisher for Earth & Environmental Science	https://www.pangaea.de/
Norwegian Marine Data Centre (NMDC)		https://data.npolar.no/home/
Bjerknes Climate Data Centre (BCDC)		https://www.bcdc.no/about.html
Norwegian Polar Data Centre		
EMSO	The European Multidisciplinary Seafloor and water column Observatory	http://emso.eu/
ICOS	The Integrated Carbon Observation System	https://www.icos-cp.eu/
HUNT Cloud, NTNU		
ENA	European Nucleotide Archive	https://www.ebi.ac.uk/ena/browser/home

NOAA NCEI	NOAA's National Centers for Environmental Information	https://www.ncei.noaa.gov/
GBIF	Global Biodiversity Information Facility	https://www.gbif.org/
PhysioNet	The Research Resource for Complex Physiologic Signals	https://physionet.org/
TROLLing	The Tromsø Repository of Language and Linguistics (TROLLing)	
NCBI	The National Center for Biotechnology Information	https://www.ncbi.nlm.nih.gov/
Viedoc		https://www.viedoc.com/
ELIXIR Core Resources and Recommended Deposition Databases.		
GitHub		
Academia.edu		
HEPData	Repository for publication-related High-Energy Physics data	https://www.hepdata.net/

VEDLEGG 2: OVERSIKT OVER SPØRSMÅL

Tema 1 - Bakgrunnsinformasjon

Nedenfor følger noen innledende spørsmål om institusjonstilhørighet og fagområde.

1.1 Hvilken type institusjon er du tilknyttet? * (mulig å krysse av flere):

- Universitet
- Høgskole
- Helseforetak
- Forskningsinstitutt
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

1.2 Hvilke(t) fagområde jobber du innen? * (mulig å krysse av flere)

- Humaniora
- Samfunnsvitenskap
- Matematikk/naturvitenskap
- Teknologi
- Medisin og helsefag
- Landbruks- og fiskerifag
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

1.3 Hvilken stilling er du ansatt i?

- Vitenskapelig assistent
- Stipendiat
- Postdoktor
- Forskerstilling
- Førsteamanuensis
- Dosent
- Professor

- Høyskole-/universitetslektor
- Førstelektor
- Ingeniør/overingeniør
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

1.4 Jobber du i forskningsgrupper eller jobber du stort sett alene? *

- Jobber stort sett alene
- Jobber stort sett i gruppe på 2-4 personer
- Jobber stort sett i en større gruppe (flere enn 4 personer)

1.5 Er du vanligvis prosjektleder/PI eller deltaker i prosjektene du deltar i? *

- Er vanligvis prosjektdeltaker
- Er vanligvis prosjektleder/PI
- Begge deler

1.6 Hvilke(n) datatype(r) håndterer du primært i dine prosjekter? * (mulig å krysse av flere)

- Data som ikke krever spesiell beskyttelse
- Persondata - ikke sensitive/særlige kategorier
- Sensitive persondata (særlige kategorier av personopplysninger)
- Andre typer data som krever beskyttelse
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

1.7 Hvordan foreligger dataene du håndterer i dine prosjekter? * (mulig å krysse av flere)

- Numeriske data (databaser, regneark, målinger, laboratoriedata, simuleringer etc.)
- Tekst (dokumenter)
- Video
- Bilde
- Lyd
- Multimedia
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

1.8 Hvor store datavolum håndterer du normalt i dine prosjekter (per prosjekt)? *

Eksempel: 25-50 timer video er ofte over 100 GB og under 1 TB

- Svært små volum (mindre enn 1 GB)
- Små volum (fra 1 GB opp til 10 GB)
- Ganske små volum (større enn 10 GB og opp til 100 GB)
- Mellomstore volum (større enn 100 GB og opp til 1 TB)
- Store volum (større enn 1 TB og opp til 10 TB)
- Svært store volum (over 10 TB)
- Vet ikke/usikker på mengde

Tema 2 – Prosjektplanlegging

I denne delen av undersøkelsen stiller vi spørsmål knyttet til den innledende fasen av forskningsprosjektene der datahåndtering planlegges.

2.1 Hvilke retningslinjer og krav til datahåndtering forholder du deg til når du planlegger dine forskningsprosjekter? * (mulig å krysse av flere)

- Retningslinjer fra institusjonen
- Nasjonale retningslinjer
- Retningslinjer fra finansiør (f.eks. Forskningsrådet eller EU)
- Retningslinjer tilknyttet infrastruktur som prosjektet benytter
- Ikke relevant - er ikke ansvarlig for planlegging av datahåndtering
- Ikke relevant - kjenner ikke til retningslinjer på området
- Ikke relevant – finnes ikke egne retningslinjer på mitt område
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

2.2 Hvilke tjenester har du brukt/bruker du under planleggingen av datahåndtering i dine forskningsprosjekter? * (mulig å krysse av flere)

- Dokumentmal fra egen institusjon
- NSD DMP
- EasyDMP (Uninett Sigma2)
- ARGOS (OpenAIRE)
- CEDAR (Stanford University)
- Data Stewardship Wizard (ELIXIR)
- The ELIXIR Research Data Management Kit (RDMkit)
- DMPonline (DCC)
- DMPTool (UC3/CDL)
- Andre
- Har ikke brukt tjenester for å lage datahåndteringsplan

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

Hvis ingen:

Hva er årsaken til at du ikke har benyttet tjenester for å lage datahåndteringsplan? * (mulig å krysse av flere)

- Jeg har ikke hatt behov for datahåndteringsplan i mine prosjekter
- Opplever at egen kunnskap om temaet er solid/tilstrekkelig
- Kolleger/samarbeidspartnere har tatt et hovedansvar for datahåndteringsplanen
- Kjenner ikke til tjenestene
- Tjenestene dekker ikke mine behov
- Annet

2.3 Hvilke øvrige veiledningstjenester har du brukt/bruker du under planleggingen av datahåndtering i dine forskningsprosjekter? * (mulig å krysse av flere)

- Kurs fra egen institusjon
- Veiledning fra forskerstøtte ved egen institusjon
- Veiledning fra kolleger ved egen institusjon
- Kurs fra nasjonal tjeneste (f.eks. NSD)
- Veiledning fra nasjonal tjeneste (f.eks. NSD eller Sigma 2)
- Andre
- Har ikke brukt øvrige veiledningstjenester

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

Hvis ingen:

Hva er årsaken til at du ikke har benyttet øvrige veiledningstjenester for planlegging av datahåndtering? * (mulig å krysse av flere)

- Jeg har ikke hatt behov for andre veiledningstjenester i mine prosjekter
- Opplever at egen kunnskap om temaet er solid/tilstrekkelig
- Kolleger/samarbeidspartnere har tatt et hovedansvar for datahåndteringsplanen
- Kjenner ikke til tjenestene
- Tjenestene dekker ikke mine behov
- Annet

2.4 Har du behov knyttet til planlegging av datahåndtering som ikke dekkes av tjenestene per i dag? (mulig å krysse av flere)

- Tjenestene dekker mine behov
- Behov for verktøy mer tilpasset mitt fagområde
- Behov for verktøy mer egnet for forskningsmetodene jeg bruker
- Behov for verktøy mer tilpasset min institusjon
- Behov for større individuell tilpasning i verktøyene
- Behov for tydeligere retningslinjer
- Behov for mer individuell veiledning
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

Tema 3 – Prosjektgjennomføring

Nedenfor følger spørsmål om den aktive fasen av forskningsprosjektene, der data samles inn, lagres og bearbeides.

3.1 Hvor lagrer du primært data i dine pågående forskningsprosjekter?* (mulig å krysse av flere)

- På egen maskin/på eget personlig område
- På ekstern harddisk/minnepinne
- I en infrastruktur/tjeneste som tilhører institusjonen
- I en nasjonal fagspesifikk infrastruktur/tjeneste
- I en nasjonal generell infrastruktur/tjeneste
- I en internasjonal fagspesifikk infrastruktur/tjeneste
- I en internasjonal generell infrastruktur/tjeneste
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

3.2 Hvilke(n) tjeneste(r) har du brukt/bruker du for å lagre data i den aktive fasen av forskningsprosjekter? * (mulig å krysse av flere)

- NIRD Project Storage (Uninett Sigma2)
- TSD (Tjenester for Sensitive Data ved UiO)
- SAFE (Sikker Adgang til Forskningsdata og E-infrastruktur ved UiB)
- OSF (Open Science Framework, Center for Open Science)
- ResearchData (lagringsplass for (større) forskningsdata ved UiT)

- Andre (f.eks. lokal tjeneste fra egen institusjon eller en fagspesifikk tjeneste)

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

3.3 Hvilke(n) tjeneste(r) har du brukt/bruker du for å analysere og bearbeide data i den aktive fasen av forskningsprosjekter? * (mulig å krysse av flere)

- NIRD Service Platform (Uninett Sigma 2)
- High Performance Computing (HPC) (Uninett Sigma 2)
- TSD (Tjenester for Sensitive Data ved UiO)
- Andre, (f.eks lokale tjenester eller internasjonale tjenester)

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

3.4 Har du behov knyttet til datahåndtering i den aktive fasen som ikke dekkes av tjenestene per i dag? * (mulig å krysse av flere)

- Tjenestene dekker mine behov
- Tjenester som gjør det enklere å dele data i den aktive fasen med kolleger ved institusjonen
- Tjenester som muliggjør deling av data i den aktive fasen med samarbeidspartnere utenfor institusjonen (i Norge og innen EU/EØS)
- Tjenester som muliggjør deling av data i den aktive fasen med samarbeidspartnere utenfor EU/EØS (definert som «tredjeland» i GDPR)
- Tjenester som veileder i å legge dokumentasjon til dataene (f.eks. legge på gode metadata)
- Tjenester som muliggjør analyse og bearbeiding av data
- Tjenester som er compatible med annen programvare jeg bruker
- Tjenester som muliggjør aggregering av data fra flere kilder
- Veiledning i valg av tjeneste/infrastruktur for lagring av data - hva kan brukes når?
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

Tema 4 – Prosjektavslutning

Nedenfor følger spørsmål knyttet til avslutningen av et forskningsprosjekt og videre arkivering og eventuelt deling av data etter prosjektets slutt.

4.1 Hvordan går du frem for å avgjøre hvilke data du skal ta vare på etter prosjektets slutt?

* (mulig å krysse av flere)

- Avgjør selv
- Diskuterer med andre forskere/leder
- Følger krav fra finansiør
- Følger krav/retningslinjer fra tidsskrift
- Følger retningslinjer fra egen institusjon
- Søker råd fra forskningsstøtte ved egen institusjon
- Søker veiledning fra nasjonal tjeneste
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

4.2a Hvilke av de følgende tjenestene har du brukt/bruker du for å arkivere eller dele data i avslutningsfasen/etter avsluttet forskningsprosjekt? * (mulig å krysse av flere)

- NIRD Research Data Archive (Uninett Sigma2)
- NSDs arkivtjenester
- TSD (Tjenester for Sensitive Data ved UiO)
- DataverseNO
- Dryad
- Harvard Dataverse
- Mendeley Data
- OSF (Open Science Framework, Center for Open Science)
- Zenodo
- Figshare
- Kaggle
- Tidsskriftspesifikk arkivering
- Andre (f.eks fagspesifikke arkiver)
- Har ikke brukt tjenester i avslutningsfasen av forskningsprosjekter

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

Hvis ingen:

Hva er årsaken til at du ikke har benyttet tjenester i avslutningsfasen/etter endt forskningsprosess? * (mulig å krysse av flere)

- Tjenestene gir ikke god nok støtte for å dele data
- Det er tidkrevende å tilrettelegge data for arkivering
- Kjenner ikke til tjenestene
- Tjenestene dekker ikke mine behov
- Annet

4.3 Har du behov knyttet til datahåndtering i avslutningsfasen som ikke dekkes av tjenestene per i dag? * (mulig å krysse av flere)

- Tjenestene dekker mine behov
- Behov for veiledning i valg av lisenser
- Behov for tjenester knyttet til å legge på dokumentasjon/metadata
- Behov for veiledning i å tilrettelegge data (valg av filnavn, formater, anonymisering etc.)
- Behov for informasjon og bistand til å gjøre dataene siterbare
- Behov for verktøy for å anonymisere data slik at de kan deles
- Annet

Om du har oppgitt annet, vær vennlig å spesifisere..
(boks til innfylling)

Tema 5 – Annet

5.1 Kunne du tenke deg å bli kontaktet med flere spørsmål om temaet? * (mulig å krysse av flere)

- Ja, jeg er villig til å besvare et spørreskjema med oppfølgingsspørsmål
- Ja, jeg er villig til å delta i fokusgrupper/intervju om temaet dersom aktuelt
- Nei takk

(Felt for å skrive inn epostadresse/kontaktinformasjon)